

# Approaching Human-Level Data Coding? A Systematic Comparison of OpenAI, DeepSeek, and Human Coders in Qualitative Analysis of Customer Reviews

Yucong Liu  
Shippensburg University

*The rapid growth of online reviews presents both opportunities and challenges for qualitative research. Human coding ensures contextual accuracy but is difficult to scale. This study compares human coding with AI-assisted coding using OpenAI GPT-4o and DeepSeek r1 on customer reviews of a complex DIY product. Results show both platforms capture underlying relationships, with OpenAI aligning more closely with human coding and DeepSeek demonstrating stronger internal consistency. Systematic AI errors mainly take the form of conservative Type I errors. Findings suggest AI can complement, rather than replace, human coders to enhance scalability and efficiency.*

*Keywords: artificial intelligence, qualitative data coding, content analysis, customer reviews, perceived helpfulness, DIY product, informational support*

## INTRODUCTION

One defining feature of the big data era is the rapid growth of text-based information, such as social media posts (e.g., meta.com, x.com) and customer reviews on e-commerce platforms (e.g., amazon.com, booking.com). To interpret such data, researchers often rely on qualitative analysis. Yet, coding qualitative data remains both challenging and time-intensive for human coders.

The rise of artificial intelligence (AI) has led many to believe that machines can now “think” and “reason” in ways similar to humans. With immense computational power and vast memory capacity, AI performs tasks once regarded as uniquely human—such as writing, analyzing data, and even generating code. Emerging platforms like OpenAI and DeepSeek have showcased extraordinary capabilities, fueling predictions that AI will fundamentally reshape work and daily life. This naturally raises the question: can AI be used for qualitative data coding?

Despite its strengths, AI faces notable limitations. Neural network-based methods are often criticized as “black boxes” (Chowdhury & Sadek, 2012). While humans supply training data and expected outputs, the AI independently determines the mapping from input to output, leaving its reasoning opaque. In other words, we cannot fully explain how AI reaches its conclusions. Beyond this concern, two additional limitations persist: (1) there is no guarantee of optimal solutions, and (2) there is no clear guidance for selecting parameters to achieve such solutions (Chowdhury & Sadek, 2012).

These strengths and limitations create a dilemma. On one hand, AI is unlikely to achieve the same level of accuracy as human coders—at least when human coding is treated as the benchmark. Unlike programming tasks, which have objective evaluation metrics, qualitative coding lacks universal standards

and is highly context dependent, making it a more demanding test of AI's capabilities. On the other hand, AI's success in writing and dismissed. One potential solution involves training AI on highly relevant datasets and providing contextual parameters for qualitative coding—yet guidance on how to set such parameters remains unclear.

In this study, we pursued a different approach. Rather than prioritizing accuracy, we examined correlations between human coding and AI coding and assessed whether AI-generated codes preserve relationships among variables. Our premise is that while AI may introduce systematic errors, those errors could remain consistent, exerting limited influence on relational patterns. If so, AI-coded data could still capture the direction and significance of relationships, even if imperfectly aligned with human coding. Specifically, we evaluated whether AI coding introduces distortions such as Type I (false negative) or Type II (false positive) errors.

To test these assumptions, we conducted a case study using Amazon.com reviews of a DIY product. A single product was selected to ensure consistency of context while maximizing variability across coding categories. Our findings show that although AI coding differs from human coding, the results are strongly correlated. Both OpenAI 4o and DeepSeek r1 produced statistically significant relationships among variables ( $p = .014$  and  $p = .059$ , respectively), compared with human coding ( $p = .005$ ). Moreover, the results suggest a tendency toward Type I error—namely, underestimation—rather than overestimation.

These findings carry important implications for research efficiency in the big data era. Qualitative coding is inherently resource-intensive: in our study, coders required extensive preparation, including literature review, contextual immersion (installation of the product), and several hours of training before completing the task in about a week. In contrast, the AI platforms completed coding in less than an hour. If AI can produce comparable results, the potential savings in time and cost are substantial.

The observed bias toward Type I error is also notable. By underestimating rather than exaggerating relationships, AI coding appears conservative—reducing the risk of overstated findings and making it more acceptable in research contexts. We view this study as a first step toward understanding how AI can complement human coders in qualitative analysis. Beyond efficiency, our results point to opportunities for integrating AI with human oversight to enhance coding accuracy. We hope this work encourages further exploration of AI's role in qualitative research.

## RESEARCH BACKGROUND

### Qualitative Data Analysis

Qualitative data analysis can be conducted at three levels: keyword mapping, content analysis, and grounded theory. Keyword mapping identifies the frequency of words or phrases (Y. Liu, 2017; Y. Liu et al., 2013, 2020; Y. Liu & Chen, 2012). Content analysis reduces text into codes that represent concepts and enables hypothesis testing about relationships among those concepts (Bernard & Ryan, 1998). Grounded theory extends further by discovering new concepts and developing theoretical models (Oktay, 2012).

Keyword mapping has already been embedded into software packages (e.g., natural language processing, NLP), and e-retailers such as Amazon.com now provide AI-generated summaries of customer reviews. By contrast, the role of AI in content analysis remains underexplored, and this is the focus of our study.

Content analysis treats text as a window into human experience, aiming to capture behaviors and experiences within conceptual categories (Bernard & Ryan, 1998). Its two core components are sampling and coding. For sampling, even a single case study may be sufficient if it captures maximum variety across variables or represents a typical phenomenon (Hartley, 2004). For coding, the primary challenge lies in ensuring that coders make accurate and consistent judgments about texts (Bernard et al., 2016).

In the big data era, vast amounts of text are available from sources such as social media. We chose to analyze customer reviews from e-commerce platforms for two reasons. First, reviews play a critical role in shaping e-commerce outcomes (Choi & Leon, 2020; Eslami et al., 2018; Ngo-Ye & Sinha, 2014), making our findings practically significant. Second, the extensive research on electronic word of mouth (eWOM) provides a rich foundation for benchmarking our results against prior studies using statistical models such

as OLS regression. Because our interest extends beyond coding accuracy to how coding influences relationships among variables, a content-analysis approach within customer review data offers an appropriate context for this research.

### **Variable Selection**

A central concept in eWOM research is *perceived helpfulness*, defined as the extent to which a customer review provides useful information for decision-making (Eslami et al., 2018). Perceived helpfulness reflects the overall quality of reviews, and prior studies have examined determinants such as review length (Chen et al., 2024; Karimi & Wang, 2017; Srivastava & Kalro, 2019), product type (Baek et al., 2012; Siering et al., 2018; Ullah et al., 2016), product rating (Deng et al., 2020; Pan & Zhang, 2011; Xu et al., 2023), and rating dispersion (Choi & Leon, 2020; S. Lee et al., 2021).

In this study, we focus on *informational support*, defined as the provision of recommendations, advice, or knowledge intended to solve problems (Ahmad & Laroche, 2017; Hajli, 2014). Informational support generally consists of three dimensions: providing information, identifying concerns, and offering solutions (Hajli, 2014; Liang et al., 2011). We selected this determinant for two reasons. First, its multidimensional structure enables a systematic comparison between AI-generated coding and human coding across distinct dimensions, allowing us to assess whether AI introduces systematic errors. Second, informational support has been validated in prior empirical studies (Ballantine & Stephenson, 2011; Hajli, 2014; Liang et al., 2011), providing a solid foundation for our analysis.

Consistent with prior findings, we expect informational support to exhibit a positive relationship with perceived helpfulness.

### **Case Selection**

We employed a case study approach to ensure contextual consistency and a deeper understanding of textual meanings. Following the guideline that a case study should capture maximum variety across variables (Bernard & Ryan, 1998), we selected a DIY product—the “360 Degree Bird View System” on Amazon.com—for three reasons. First, installing this product often requires many hours or even days, creating a strong need for informational support in the form of providing information, identifying concerns, and offering solutions. Second, the seller offered limited installation guidance, which further heightened customers’ reliance on peer-generated informational support. Third, the product remained available on Amazon for five years (2019–2023), suggesting that customer reviews provided sufficient support for overcoming installation challenges. Taken together, these factors make this product an appropriate and robust case for our research purpose.

### **Data Coding**

The human coding process was carefully designed to ensure accurate interpretation of installation issues. To gain firsthand experience, the authors purchased the product twice and installed it on two different vehicles. This background knowledge enhanced confidence in identifying information, concerns, and solutions expressed in customer reviews. Three undergraduate students with DIY automotive experience were recruited as coders. They were blind to the study’s purpose and hypotheses and were tasked solely with coding.

Coder training consisted of three steps: (1) an introduction to the product and its installation process, (2) instruction on the definition and three dimensions of informational support, followed by open discussion, and (3) a pilot coding exercise with 20 reviews, jointly reviewed to resolve discrepancies and achieve convergence.

Following training, coders independently coded all reviews. Because the sample size was relatively small to avoid helpful-voting bias, i.e., customers voting on only a subset of reviews (J. Liu et al., 2007), we employed a fully crossed design (Hallgren, 2012), where all coders coded all data. Inter-rater reliability in the first round was 0.82, indicating strong agreement (Aguirre-Urreta et al., 2010; Ellis et al., 2016; Y. Lee et al., 2007). Discrepant cases were discussed in a second round until full consensus was reached.

For AI-assisted coding, we employed OpenAI GPT-4o (chatgpt.com) and DeepSeek r1 (chat.deepseek.com).

## DATA ANALYSIS AND COMPARISON

### Descriptions, Trends and Corrections of Dimensions

Table 1 presents the descriptive statistics of the coded variables. The means and standard deviations of the OpenAI-generated data exhibit greater similarity to those of the human-coded data compared to the outputs from DeepSeek. This pattern suggests that OpenAI’s algorithmic processes may be more closely aligned with human judgment, whereas DeepSeek appears to rely on different computational mechanisms. These differences have important implications for the validity and robustness of AI-assisted qualitative coding, highlighting the need to carefully evaluate platform-specific outputs before drawing research conclusions.

**TABLE 1**  
**DESCRIPTIONS OF DATA**

Dimensions	minimum	maximum	mean	Std. Deviation
1-PI	0	9	.8209	1.78313
1-IC	0	7	1.4478	1.56944
1-OS	0	8	.9254	1.56004
2-PI	0	5	.8060	1.18371
2-IC	0	6	1.7424	1.64841
2-OS	0	4	.4776	.94321
3-PI	1	12	3.5075	2.10607
3-IC	0	8	2.6716	1.96475
3-OS	0	6	2.3433	1.48275

1: human 2. OpenAI 3. DeepSeek

PI: providing information IC: identifying concerns OS: offering solutions

Figure 1 illustrates the trends across the three dimensions. While the coded data are not identical in most cases, they exhibit strong correlations. Notably, the OpenAI-generated data align more closely with the human-coded data than those produced by DeepSeek. For this case study, this suggests that OpenAI provides a closer approximation to human coding, whereas DeepSeek introduces greater variability in capturing informational support.

**FIGURE 1**  
**TRENDS OF THREE**

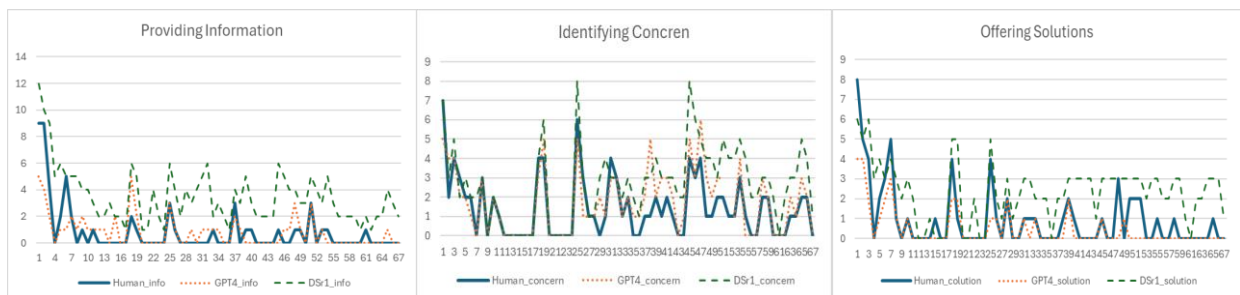


Table 2 presents the correlations among variables. Across the three sources, the dimensions are highly correlated and statistically significant at the 0.01 level (two-tailed). For the providing information

dimension, the correlation coefficients between human and OpenAI, and between human and DeepSeek, are .730 and .783, respectively, while the coefficient between OpenAI and DeepSeek is .703. For the identifying concerns dimension, the coefficients are .825 (human–OpenAI), .790 (human–DeepSeek), and .814 (OpenAI–DeepSeek). For the offering solutions dimension, the coefficients are .838 (human–OpenAI), .692 (human–DeepSeek), and .639 (OpenAI–DeepSeek). Overall, the correlations between human and OpenAI are generally higher than those between human and DeepSeek, reinforcing the pattern observed in Figure 1. This indicates that, within the context of this case study, OpenAI captures relational patterns in customer reviews more consistently with human coders than DeepSeek.

**TABLE 2  
CORRECTION MATRIX**

	1-PI	1-IC	1-OS	2-PI	2-IC	2-OS	3-PI	3-IC
1-IC	.478**							
1-OS	.801**	.571**						
2-PI	.730**	.529**	.755**					
2-IC	.386**	.825**	.444**	.455**				
2-OS	.781**	.508**	.838**	.627**	.357**			
3-PI	.783**	.760**	.796**	.703**	.640**	.761**		
3-IC	.359**	.790**	.442**	.396**	.814**	.331**	.608**	
3-OS	.562**	.760**	.692**	.548**	.699**	.639**	.788**	.757**

\*\* Correlation is significant at the 0.01 level (2-tailed).

1: Human 2. OpenAI 3. DeepSeek

PI: Providing information IC: Identifying concerns OS: Offering solutions

### Factor Analyses

The multidimensional structure of informational support enabled us to conduct factor analyses to examine how the three dimensions relate to the underlying latent construct. All factor loadings were significant, with the lowest at .717. The composite reliabilities (CRs) were 0.899 for human coding, 0.850 for OpenAI coding, and 0.928 for DeepSeek coding. The average variances extracted (AVEs) were 0.748 (human coding), 0.656 (OpenAI coding), and 0.812 (DeepSeek coding). All values exceeded the conventional thresholds (factor loading > 0.70, CR > 0.80, AVE > 0.50) (Fornell & Larcker, 1981; Y. Liu et al., 2011; Y. Liu & Chen, 2008).

These results indicate that AI algorithms demonstrate consistency when applied to qualitative data coding. In this case study, OpenAI more closely approximated human coding overall (as shown in Table 2 and Figure 1), while DeepSeek exhibited stronger internal reliability across the three dimensions (AVE = 0.812 vs. 0.656 for OpenAI). Taken together, these findings suggest that the two platforms offer complementary strengths: OpenAI excels in external alignment with human judgment, whereas DeepSeek provides higher internal consistency in representing the multidimensional construct.

**TABLE 3**  
**FACTOR LOADINGS, COMPOSITE RELIABILITY (CR), AND AVERAGE VARIANCE EXTRACTED (AVE)**

Components	Loadings	Composite Reliability (CR)	Average Variance Extracted (AVE)
Human-PI	.892	0.899	0.748
Human-IC	.927		
Human-OS	.768		
OpenAI-PI	.874	0.850	0.656
OpenAI-IC	.717		
OpenAI-OS	.830		
DeepSeek-PI	.886	0.928	0.812
DeepSeek-IC	.872		
DeepSeek-OS	.944		

**OLS Regression Analysis**

Table 4 reports the results of the regression analyses using the three data sets. Human coding produced the strongest results, with the largest coefficient (.339), the highest explanatory power (adjusted R<sup>2</sup> = .101), and the most significant p-value (p < .001). OpenAI-4o coding followed, with a coefficient of .301, an adjusted R<sup>2</sup> of .077, and significance at the 0.05 level. DeepSeek-r1 coding showed the weakest results, with a coefficient of .232, an adjusted R<sup>2</sup> of .039, and marginal significance at the 0.10 level. These findings confirm the role of human coding as the benchmark for comparison, while also demonstrating that OpenAI-4o approximates human coding more closely than DeepSeek-r1 in this case study. Importantly, this hierarchy suggests that while AI-based coding can be used to support regression-based eWOM research, its reliability and explanatory power vary across platforms, underscoring the need for platform-specific validation.

**TABLE 4**  
**RESULTS OF REGRESSION ANALYSIS (DEPENDENT VARIABLE: HELPFULNESS)**

	Constant	Standardized Coefficients	P-Value	Adjusted R Square
Human Informational Support	1.183***	.339***	.005	.101
OpenAI4o Informational Support	1.149***	.301**	.014	.077
DeepSeekr1 Informational Support	0.999***	.232*	.059	.039

\*\*\*: p-value < 0.01 \*\*: p-value < 0.05 \*: p-value < 0.1

**CONCLUSION AND IMPLICATIONS**

The results of our analyses support the argument that the shifts observed in AI-coded data relative to human coding reflect systematic errors. Given these errors, achieving high levels of accuracy with AI in qualitative data coding does not appear feasible. Nevertheless, both AI platforms were able to capture the underlying relationships among variables, with OpenAI producing results that more closely approximate human coding and DeepSeek yielding acceptable outcomes. Importantly, these systematic errors primarily manifest as Type I errors—that is, conservative underestimations—rather than overestimations, which is generally preferable in research contexts.

An additional contribution of this study lies in revealing platform-specific differences. OpenAI demonstrated stronger external validity by aligning more closely with human-coded results, whereas DeepSeek exhibited higher internal consistency across dimensions of informational support. These complementary strengths suggest that platform choice may depend on research priorities: OpenAI may be better suited when the goal is to replicate human judgments, while DeepSeek may be preferable when reliability within multidimensional constructs is emphasized.

These findings have several important implications. First, accuracy has never been the defining characteristic of big data analytics. By the same logic, AI-generated coding results cannot guarantee optimal solutions, and thus pursuing perfect accuracy is not a practical goal. Instead, our study suggests shifting the focus from accuracy to correlation. If AI-generated coding is highly correlated with human coding, even without perfect alignment, the results of subsequent data analyses can still be considered valid.

Second, our findings indicate that AI can be effectively employed for qualitative data coding, albeit with conservative tendencies. Given the overwhelming scale of text data in the big data era, the primary limitation of qualitative research lies in the human inability to code data at such volume. If AI can assume this role, researchers will be able to conduct qualitative studies more efficiently and at scale, while still producing reliable outcomes.

Finally, in contexts where coding accuracy remains a priority, AI-based coding can still serve as a useful baseline for human coders. Moreover, we expect that AI-based coding may help guide researchers in setting parameters and developing strategies to further improve coding quality. Taken together, our results suggest that AI should not be viewed as a replacement for human coders but rather as a complementary tool that can enhance the efficiency, scalability, and robustness of qualitative research.

## **LIMITATIONS AND FUTURE RESEARCH**

As with any case study, the findings of this research should be interpreted with caution. Our analysis was limited to a single product category (a DIY automotive installation system) and a relatively small sample size. While this context provided the necessary depth to examine the nuances of informational support in customer reviews, it also restricts the generalizability of our results. Future research should validate these findings across a broader range of products and larger datasets to assess the robustness of the observed patterns. In particular, examining products that vary in complexity, usage frequency, or consumer involvement would provide greater insights into the applicability of AI for qualitative data coding.

A second limitation lies in the reliance on two AI platforms, OpenAI and DeepSeek, which differ in their underlying algorithms and performance characteristics. Although our study highlights complementary strengths—OpenAI aligning more closely with human coding and DeepSeek demonstrating higher internal reliability—further research is needed to systematically evaluate how different AI models compare in terms of accuracy, consistency, and error patterns. Future studies could also explore whether combining multiple AI outputs, or integrating AI with human oversight, produces superior results.

Finally, while we identified systematic errors in AI coding, particularly conservative Type I errors, further research should delve deeper into understanding the sources of these discrepancies. Identifying where and why AI diverges from human judgment could inform targeted training strategies to improve AI's ability to capture qualitative nuances. Such efforts may lead to the development of hybrid coding frameworks, in which AI handles large-scale data processing while human coders focus on complex or ambiguous cases. This approach may offer a practical path forward for enhancing the efficiency and accuracy of qualitative research in the big data era.

## REFERENCES

- Aguirre-Urreta, M.I., Ellis, M.E., Sun, W.N., Liu, Y., Lee, K.K., Mao, J., & Marakas, G.M. (2010). How many technology types are there? Preliminary results from the technology acceptance literature. *AMCIS 2010 Proceedings*. Americas Conference on Information Systems.
- Ahmad, S.N., & Laroche, M. (2017). Analyzing electronic word of mouth: A social commerce construct. *International Journal of Information Management*, 37(3), 202–213. <https://doi.org/10.1016/j.ijinfomgt.2016.08.004>
- Baek, H., Ahn, J., & Choi, Y. (2012). Helpfulness of online consumer reviews: Readers' objectives and review cues. *International Journal of Electronic Commerce*, 17(2), 99–126.
- Ballantine, P.W., & Stephenson, R.J. (2011). Help me, I'm fat! Social support in online weight loss networks. *Journal of Consumer Behaviour*, 10(6), 332–337. <https://doi.org/10.1002/cb.374>
- Bernard, H.R., & Ryan, G. (1998). Text analysis. *Handbook of Methods in Cultural Anthropology*, 613.
- Bernard, H.R., Wutich, A., & Ryan, G.W. (2016). *Analyzing qualitative data: Systematic approaches*. SAGE publications.
- Chen, K., Tsai, C.-F., Hu, Y.-H., & Hu, C.-W. (2024). The effect of review visibility and diagnosticity on review helpfulness – An accessibility-diagnostics theory perspective. *Decision Support Systems*, 178, 114145. <https://doi.org/10.1016/j.dss.2023.114145>
- Choi, H.S., & Leon, S. (2020). An empirical investigation of online review helpfulness: A big data perspective. *Decision Support Systems*, 139, 113403.
- Chowdhury, M., & Sadek, A.W. (2012). Advantages and limitations of artificial intelligence. *Artificial Intelligence Applications to Critical Transportation Issues*, 6(3), 360–375.
- Deng, W., Yi, M., & Lu, Y. (2020). Vote or not? How various information cues affect helpfulness voting of online reviews. *Online Information Review*, 44(4), 787–803.
- Ellis, M.E., Aguirre-Urreta, M.I., Lee, K.K., Sun, W.N., Liu, Y., Mao, J., & Marakas, G.M. (2016). Categorization of technologies: Insights from the technology acceptance literature. *Journal of Applied Business and Economics*, 18(4), 20–30.
- Eslami, S.P., Ghasemaghaei, M., & Hassanein, K. (2018). Which online reviews do consumers find most helpful? A multi-method investigation. *Decision Support Systems*, 113, 32–42. <https://doi.org/10.1016/j.dss.2018.06.012>
- Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Hajli, M.N. (2014). The role of social support on relationship quality and social commerce. *Technological Forecasting and Social Change*, 87, 17–27. <https://doi.org/10.1016/j.techfore.2014.05.012>
- Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23.
- Hartley, J. (2004). What is a case study. *Essential Guide to Qualitative Methods in Organizational Research*, 323.
- Karimi, S., & Wang, F. (2017). Online review helpfulness: Impact of reviewer profile image. *Decision Support Systems*, 96, 39–48. <https://doi.org/10.1016/j.dss.2017.02.001>
- Lee, S., Lee, S., & Baek, H. (2021). Does the dispersion of online review ratings affect review helpfulness? *Computers in Human Behavior*, 117, 106670.
- Lee, Y., Lee, J.-Y., & Liu, Y. (2007). Protection motivation theory in information system adoption: A case of anti-plagiarism system. *AMCIS 2007 Proceedings*, 62.
- Liang, T.-P., Ho, Y.-T., Li, Y.-W., & Turban, E. (2011). What drives social commerce: The role of social support and relationship quality. *International Journal of Electronic Commerce*, 16(2), 69–90.
- Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 334–342.

- Liu, Y. (2017). The Moderating Effects of Network Centrality between IT Initiatives and Firm Performance: An Empirical Study. *GSTF Journal on Computing (JoC)*, 5(2), 63.
- Liu, Y., & Chen, A. (2012). *Business Value of Information Technology in A Network Environment*. Unpublished doctoral dissertation. University of Kansas.
- Liu, Y., & Chen, A.N. (2008). The effect of individual differences, tasks, and decision models on user acceptance of decision support systems. *AMCIS 2008 Proceedings*, 357.
- Liu, Y., Lee, Y., & Chen, A.N. (2011). Evaluating the effects of task–individual–technology fit in multi-DSS models context: A two-phase view. *Decision Support Systems*, 51(3), 688–700.
- Liu, Y., Lee, Y., & Chen, A.N. (2013). Information Technology and Firm Profitability in Network Environments. *AMCIS 2013 Proceedings*. Americas Conference on Information Systems, Chicago, Illinois, USA.
- Liu, Y., Lee, Y., & Chen, A.N. (2020). How IT wisdom affects firm performance: An empirical investigation of 15-year US panel data. *Decision Support Systems*, 133, 113300.
- Ngo-Ye, T.L., & Sinha, A.P. (2014). The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61, 47–58. <https://doi.org/10.1016/j.dss.2014.01.011>
- Oktay, J.S. (2012). *Grounded theory*. Oxford University Press.
- Pan, Y., & Zhang, J.Q. (2011). Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews. *Journal of Retailing*, 87(4), 598–612. <https://doi.org/10.1016/j.jretai.2011.05.002>
- Siering, M., Muntermann, J., & Rajagopalan, B. (2018). Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, 108, 1–12. <https://doi.org/10.1016/j.dss.2018.01.004>
- Srivastava, V., & Kalro, A.D. (2019). Enhancing the helpfulness of online consumer reviews: The role of latent (content) factors. *Journal of Interactive Marketing*, 48(1), 33–50.
- Ullah, R., Amblee, N., Kim, W., & Lee, H. (2016). From valence to emotions: Exploring the distribution of emotions in online product reviews. *Decision Support Systems*, 81, 41–53. <https://doi.org/10.1016/j.dss.2015.10.007>
- Xu, C., Zheng, X., & Yang, F. (2023). Examining the effects of negative emotions on review helpfulness: The moderating role of product price. *Computers in Human Behavior*, 139, 107501. <https://doi.org/10.1016/j.chb.2022.107501>